



University of Pennsylvania
Scholarly Commons

Marketing Papers

11-21-2011

Illusions in Regression Analysis

J. Scott Armstrong

University of Pennsylvania, armstrong@wharton.upenn.edu

Armstrong, J.S. (2011). Illusions in Regression Analysis. Forthcoming in *International Journal of Forecasting*, 2012.

This paper is posted at Scholarly Commons. http://repository.upenn.edu/marketing_papers/173
For more information, please contact repository@pobox.upenn.edu.

Illusions in Regression Analysis

Abstract

Soyer and Hogarth's article, "The Illusion of Predictability," shows that diagnostic statistics that are commonly provided with regression analysis lead to confusion, reduced accuracy, and overconfidence. Even highly competent researchers are subject to these problems. This overview examines the Soyer-Hogarth findings in light of prior research on illusions associated with regression analysis. It also summarizes solutions that have been proposed over the past century. These solutions would enhance the value of regression analysis.

Disciplines

Business | Social and Behavioral Sciences

Comments

Armstrong, J.S. (2011). Illusions in Regression Analysis. Forthcoming in *International Journal of Forecasting*, 2012.

Illusions in Regression Analysis

J. Scott Armstrong (armstrong@wharton.upenn.edu)
The Wharton School, University of Pennsylvania, Philadelphia PA 19104
Forthcoming, *International Journal of Forecasting*, 2012

Abstract

Soyer and Hogarth's article, "The Illusion of Predictability," shows that diagnostic statistics that are commonly provided with regression analysis lead to confusion, reduced accuracy, and overconfidence. Even highly competent researchers are subject to these problems. This overview examines the Soyer-Hogarth findings in light of prior research on illusions associated with regression analysis. It also summarizes solutions that have been proposed over the past century. These solutions would enhance the value of regression analysis.

Keywords: *a priori* analysis, decision-making, *ex ante* testing, forecasting, non-experimental data, statistical significance, uncertainty

The "Illusion of Predictability: How Regression Statistics Mislead Experts," by Emre Soyer and Robin Hogarth, is dedicated to the memory of Arnold Zellner (1927-2010).¹ I am sure that Arnold would have agreed with me that their paper is a fitting tribute.

Given the widespread use of regression analysis, the implications of the article are important for the life and social sciences. Employing a simple experiment, Soyer and Hogarth (2011, hereafter "S&H") show that some of the world's leading experts in econometrics can be misled by standard statistics provided with regression analyses: *t*, *p*, *F*, *R*-squared and the like.

S&H follows a rich history on the illusions of predictability associated with the use of regression analysis on non-experimental data. A look at the history of regression analysis suggests why illusions of predictability occur and why they have increased over time – to the detriment, as S&H show, of scientific analysis and forecast-ability.²

Historical view of illusions in regression analysis

Regression analysis entered the social sciences in the 1870s with the pioneering work by Francis Galton. But "least squares" goes back at least to the early 1800s and the German mathematician Karl Gauss, who used the technique to predict astronomical phenomena.

For most of its history, regression analysis was a complex, cumbersome, and expensive undertaking.

¹For more about Arnold Zellner, see "In memory of Arnold Zellner, a great person and scientist," *International Journal of Forecasting*, 27 (2011), 961-967, by Antonio García-Ferrer.

²There is also an extensive literature in which leading econometricians have warned of illusions associated with regression (e.g., see the review by Kennedy, 2002). Karmi and Shapiro (1980) made fun of regression analysis procedures in their "data torture" paper.

Consider Milton Friedman's experience more than forty years prior to user-friendly software and the personal computer revolution. Around 1944, as part of the war effort, Friedman was asked to analyze data on alloys used in turbine engine blades. He used regression analysis to develop a model that predicted time to failure as a function of stress, temperature, and some metallurgical variables representing the alloy's composition. Obtaining estimates for Friedman's equation by hand and calculating test statistics would have taken a skilled analyst about three months labor. Fortunately, a large computer, built from many IBM card-sorters and housed in Harvard's air-conditioned gymnasium, could do the calculations. Ignoring time required for data input, the computer needed 40 hours to calculate the regression estimates and test statistics. Today, a regression of the size and complexity of Friedman's could be executed in about one second.

Friedman was delighted with the results; the model had a high R^2 and the variables were "statistically significant" at conventional levels. As a result, Friedman recommended two new improved alloys, which his model predicted would survive several hundred hours at high temperatures. Tests of the new alloys were carried out by engineers in an MIT laboratory. The result? The first alloy broke in about two hours and the second one in about three. Friedman concluded one should focus on tests of outputs (forecasts) rather than statistically significant inputs. He also opined that "the more complex the regression, the more skeptical I am" (Friedman and Schwartz 1991).

Given that doing regressions was expensive, it was sensible to rely heavily on *a priori* analyses. Opinions about the proper model (e.g., which variables are important) should not be based on opinions or untested ideas even though they may be offered by famous people. (The names Keynes and Samuelson spring to mind.) Instead, the evidence should be based on meta-analyses. Meta-analyses produce more accurate and less biased summaries than those provided by traditional reviews as shown by Cumming (2012).

When possible, meta-analyses should be used in making decisions as to what variables to include; specifying the expected direction of the relationships; and specifying the nature of the functional form, ranges of magnitudes of relationships, and size of expected magnitudes of those relationships. It is also important to determine relationships that can be measured outside the model based either on common knowledge (for example, adjusting for inflation or transforming the data to a per capita basis) or on analyses of other data.

Analysts should use simple pre-specified rules to combine *a priori* estimates with estimates obtained from regression analysis (for example, one might weight each estimate of a relationship equally, then re-run the regression to estimate the coefficients for the other variables). This approach was recommended by Wold and Jureen (1953), where it was called "conditional regression." I think of it as a 'poor man's Bayesian analysis.' I prefer it to formal Bayesian forecasting methods because of its clarity about the nature of each causal relationship and the related evidence (and because I have a strong need for sleep whenever I try to read a paper on Bayesian forecasting.) In this paper, I refer to it as an *a priori* analysis.

In the mid-1960s, I was working on my PhD thesis at MIT. While the cost of regression analysis had plunged, it still involved punch cards and overnight runs. But the most time-consuming part of my thesis was the *a*

a priori analysis. Before doing any regression analyses, I gave John Little, my thesis advisor, *a priori* estimates of the coefficients for all variables in a demand-forecasting model. As it turned out, these purely *a priori* models provided relatively accurate forecasts on their own. I then used regression analyses of time-series, longitudinal, and household data to estimate parameters. These were used to revise the *a priori* estimates. This procedure provided forecasts that were substantially more accurate than those from extrapolation methods and from stepwise regression on the complete set of causal variables that were considered (Armstrong 1968a,b).

Despite warnings over the past half-century or more (Zellner, 2001, traces this back to Sir Harold Jeffreys in the mid-1900s), *a priori* analysis seems to be giving way among academic researchers to the belief that with enormous databases they can use complex methods and analytical measures such as R^2 and t-statistics to *create* models. They even try various transformations or different lags of variables to see which best fit the historical data. Einhorn (1972) concluded, “Just as the alchemists were not successful in turning base metal into gold, the modern researcher cannot rely on the ‘computer’ to turn his data into meaningful and valuable scientific information.” Ord (2012) provides a simple demonstration of how standard regression procedures, applied without *a priori* analyses, can lead one astray.

Forecast accuracy and confidence

We have ample evidence that regression analysis often provides useful forecasts (Armstrong 1985; Allen and Fildes 2001). Regression-based prediction is most effective when dealing with a small number of variables, large amounts of reliable and valid data, where changes are expected to be large and predictable, and when using well-established causal relationships – such as the elasticities for income, price, and advertising when forecasting demand. However, there are illusions that reduce the forecast accuracy and lead to overconfidence in regression analysis. I discuss five of them here:

Complexity illusion: It seems common sense that complex solutions are needed for complex and uncertain problems. Research findings suggest the opposite. For example, Christ (1960) found that simultaneous equations provided forecasts that were more accurate than those from simpler regression models when tested on artificial data, but not when tested out of sample using real data. My summary of the empirical evidence concluded that increased complexity of regression models typically reduced forecast accuracy (Armstrong 1985, pp. 225-232). Zellner (2001) reached the same conclusion in his review of the research. He also found that many users have become disillusioned with complicated models. For example he reported “the Federal Reserve Bank of Minneapolis decided to scrap its complicated vector autoregressive (VAR, i.e., Very Awful Regression) models after their poor performance in forecasting turning points, etc.”

Evidence favoring simplicity has continued to appear over the past quarter century. Why then is there such a strong interest in complex regression analyses? Perhaps this is due to academics’ preference for complex solutions, as

Hogarth (2012) describes.

Somewhere I encountered the idea that statistics was supposed to aid communication. Complex regression methods and a flock of diagnostic statistics have taken us in the other direction.

The solution is that, when specifying a model, rely upon *a priori analysis*. Follow Zellner's (2001) advice and use Occam's Razor.³ In other words, keep it simple. Start with a very simple model, such as a no-change model, and then add complexity only if there is experimental evidence to support the complication. And do not try to estimate relationships for more than three variables in a regression (findings from Goldstein and Gigerenzer, 2009, are consistent with this rule-of-thumb).

Illusion that regression models are sufficient: Forecasts are often derived only from what is thought to be the best model. This belief has a long history in forecasting.

For solutions, I call your attention to two of the most important findings in forecasting. First is that the naïve or no-change model is often quite accurate. It is to forecasting what the placebo is to medicine. This approach is especially difficult to beat in situations involving complexity and uncertainty. Here, it often helps to shrink each coefficient toward having no effect (but remember to re-run the regression to calibrate the constant term).

Second is the benefit of combining forecasts. That is, find two or more valid forecasting methods and then calculate averages of their forecasts. For example, make forecasts by using different regression models, and then combine the forecasts. This is especially effective when the methods, models, and data differ substantially. Combining forecasts has reduced errors from about 10% to 58% (depending on the conditions) compared to the average errors of the uncombined individual forecasts (Graefe, et al 2011).

Illusion that regression provides the best linear unbiased estimators: Statisticians have devoted much time to showing that regression leads to the best estimates of relationships. However, studies have shown that regression estimates produce *ex ante* forecasts that are often *less accurate* than forecasts from "unit weights" models. Schmidt (1971) was one of the first to test this idea and he found that unit weights were superior to regression weights when the regressions were based on many variables and small sample sizes. Einhorn and Hogarth (1975) and Dana and Dawes (2004) show the conditions under which regression is and is not effective relative to equal weights.

One good characteristic of regression estimates is that they become more conservative as uncertainty increases. Unfortunately, some aspects of uncertainty are ignored. For example, the coefficients can "get credit" for important excluded variables that happen to be correlated with the predictor variables. Adding variables to the regression cannot solve this problem.

In addition, analysts searching for the best fit, and publication practices favoring statistically significant

³Zellner traced this idea back to Sir Harold Jeffreys in the 1930s.

results often defeat conservatism. Thus, regressions typically over-estimate change. Ioannidis (2005, 2008) provides more reasons why regressions over-estimate changes.

One solution is to combine forecasts, and among the alternative models, include the naïve model. In particular, for problems involving time-series, damp the forecasts more heavily toward the naïve model forecasts as the forecast horizon increases. This is done to reflect effects of the increasing amount of uncertainty in the more distant future.

Illusion of control: Users of regression assume that by putting variables into the equation they are somehow controlling for these variables. This only occurs for experimental data. Adding variables does not mean controlling for variables in non-experimental data because many variables typically co-vary with other predictor variables. The problem becomes worse as variables are added to the regression. Large sample sizes cannot resolve this problem, so statistics on the number of degrees of freedom are misleading.

One solution is to use evidence from experimental studies to estimate effects and then adjust the dependent variable for these effects.

“Fit implies accuracy” illusion: Analysts assume that models with a better fit provide more accurate forecasts. This ignores the research showing that fit bears little relationship to *ex ante* forecast accuracy, especially for time series. Typically, fit improves as complexity increases, while *ex ante* forecast accuracy decreases – a conclusion that Zellner (2001) traced back to Sir Harold Jeffreys in the 1930s. In addition, analysts use statistics to improve the fit of the model to the data. In one of my Tom Swift studies, Tom used standard procedures when starting with 31 observations and 30 potential variables. He used stepwise regression and included only variables where t was greater than 2.0. Along the way, he dropped three outliers. The final regression had eight variables and an R-square (adjusted for degrees of freedom) of 0.85. Not bad, considering that the data were from Rand's book of random numbers (Armstrong 1970).

I traced studies on this illusion back to at least 1956 in an early review of the research on fit and accuracy (Armstrong 1985). Studies have continued to find the fit is not a good way to assess predictive ability (e.g., Pant and Starbuck 1990).

The obvious solution is to avoid use of t , p , F , R -squared and the like when using regression.

Using regression for decision-making

Regression analysis provides an objective and systematic way to analyze data. As a result, decisions based on regression are less likely to be subject to bias, they are consistent, the basis for the decisions can be fully explained – and they are generally useful. The gains are especially well documented when compared to judgmental decisions based on the same data (Grove and Meehl 1996; Armstrong 2001). However, two illusions, statistical significance and

correlations, can reduce the value of regression analysis.

Statistical significance illusion: S&H incorporate the illusion due to tests of statistical significance. Meehl (1978) concluded that “reliance on merely refuting the null hypothesis . . . is basically unsound, poor scientific strategy, and one of the worst things that ever happened in the history of psychology

Schmidt (1996) offered the following challenge: “Can you articulate even one legitimate contribution that significance testing has made (or makes) to the research enterprise (i.e., any way in which it contributes to the development of cumulative scientific knowledge)?” One might also ask if there is a study wherein statistical significance improves decision-making. In contrast, it is easy to find cases where statistical significance harmed decision-making. Ziliak and McCloskey (2008) document in devastating examples taken from across the sciences. To offer another example, Hauer (2004) demonstrates harmful decisions related to automobile traffic safety, such as the “Right-turn-on-red decision.” Cumming (2012) describes additional examples of the harm caused by the use of statistical significance.

The commonly recommended solution is to use confidence intervals and avoid the use of statistical significance. Statisticians argue that statistical significance provides the same information as confidence intervals. But the issue is how people use the information. Significance levels lead to confusion even among leading researchers. Cumming (2012, pp. 13-14) describes an experiment showing that when researchers in psychology, behavioral neuroscience, and medicine were presented with a set of results, 40% of 55 participants who used significance levels to guide their interpretation reached correct conclusions. In stark contrast, 95% of the 57 participants who thought in terms of confidence intervals reached correct conclusions.

Correlation illusion: We all claim to understand that correlation is not causation. The correlations might occur because A causes B, or B causes A, or they are each related to C, or they could be spurious. But when presented with sophisticated and complex regressions, people often forget that; Researchers in medicine, economics, psychology, finance, marketing, sociology, and so on, fill journals and newspapers with interesting but erroneous—and even costly—findings.

In one study, we had an opportunity to compare findings from experiments with those from analyses of non-experimental data for 24 causal statements. The directional effects differed for 8 of the 24 comparisons (Armstrong and Patnaik 2009). My conclusion is that analyses of non-experimental are often misleading.

This illusion has led people to make poor decisions about such things as what to eat (e.g., coffee, once bad, is now good for health), what medical procedures to use (e.g., the frequently recommended PSA test for prostate cancer has now been shown to be harmful), and what economic policies the government should adopt in recessions (e.g., trusting the government to be more efficient than the market).

According to Zellner (2001), Sir Harold Jeffreys had warned of this illusion, and, in 1961, referred to it as

the “most fundamental fallacy of all.”

The solution is to base causality on meta-analyses of experimental studies.

Discussion

An obvious conclusion from the study by S&H is to de-emphasize descriptive statistics for regression packages. Software developers should provide statistics on the ability of alternative methods to produce accurate forecasts on holdout samples as the default option. They could allow users to click a button to access the traditional regression statistics, a warning label should be provided near that button.

S&H, echoing Friedman, emphasize that scientific theories should be tested for their predictive ability relative to other methods. Ord (2012) deplores the fact that few regression packages aid in such analyses. It would be helpful if software providers would focus on *ex ante* testing by making it easy to simulate the forecasting situation. For cross-sectional forecasts, use jackknifing—that is, use all but one data point to estimate the model, then predict for the excluded observation, and repeat until predictions have been made for each observation in the data. For time-series, withhold data, then use successive updating and report the accuracy for each forecast horizon. These testing procedures are less likely to lead to overconfidence because they include the uncertainty from errors due to overfitting and errors in forecasting the predictor variables.

Software packages should provide statistics that allow for meaningful comparisons among methods (Armstrong and Collopy 1992). The MdRAE (Median Relative Absolute Error) was designed for such comparisons and many software packages now provide this statistic, so it should be among the default statistics, along with a link to the literature on this topic. Do not provide RMSE (Root Mean Square Errors) as it is unreliable and uninformative.

Allen and Fildes (2001) note that since 1985 there has been a substantial increase in the attention paid to *ex ante* forecast comparisons in published papers. This is consistent with the aims stated in the founding of the *Journal of Forecasting* in 1982, and subsequently, the *International Journal of Forecasting*.

Regression analysis is clearly one of the most important tools available to researchers. However, it is not the only game in town. Researchers made scientific discoveries about causality prior to the availability of regression analysis as shown by Freedman (1991) in his paper aptly titled “Statistical models and shoe leather.” He demonstrates how major gains were made in epidemiology in the 1800s. For example, John Snow’s discovery of the cause of cholera in London in the 1850s came about from “the clarity of the prior reasoning, the bringing together of many different lines of evidence, and the amount of shoe leather Snow was willing to use to get the data.” These three characteristics of good science, described by Freedman, are missing from most regression analyses that I see in journals.

We would be wise to recall a method that Ben Franklin used to address the issue of how to make decisions when many variables are involved (Sparks, 1844).⁴ He suggested listing the variables related to the choice between

⁴Benjamin Franklin’s letter to Joseph Priestley can be found at: <http://www.procon.org/viewbackgroundresource.asp?resourceID=1474>.

two options, identify which option is better for each variable, weight the variables, and then add. Pick the option that has the highest score. Andreas Graefe and I have built upon Franklin's advice in developing what we call the *index method* for forecasting. The method relies only on *a priori* analysis (preferably experimental findings) to determine which variables are important and what is the direction of the effect for each variable. Franklin suggested differential weights, but the literature discussed above suggests that unit weights are a good place to start. Regression analyses can then be used to estimate the effects of an index score. The index model allows analysts to take account of "the knowledge present in a field" as recommended by Zellner (2001). The few tests to date suggest that the index method provides useful forecasts when there are many important variables and substantial prior knowledge (e.g., see Armstrong and Graefe 2011).

It can be expensive to do *a priori* analyses for complex situations. For example, I am currently involved in a project to forecast the effectiveness of advertisements by using the index method. I have spent many hours over a 16-year period summarizing the knowledge. This led to 195 principles (causal condition/action statements), each based on meta-analysis when there was more than one source of evidence. The vast majority of them were sufficiently complex such that neither prior experience nor regression analyses were able to discover them. They were formulated thanks to a century of experimental and non-experimental studies. None of these evidence-based principles were found in the advertising textbooks and handbooks that I analyzed. Many are counter-intuitive and are often violated by advertisers (Armstrong 2011). Early findings suggest that the index model provides useful forecasts in this situation. In contrast, regression analyses have met repeated failures in this area because there may be over 50 principles that were used – or misused – in an ad.

As S&H suggest, further experimentation is needed. We need experiments to assess the ability of alternative techniques to improve accuracy when tested on large samples of forecasts on holdout samples. Journal editors should commission such studies. Allen and Fildes (2001) provide an obvious starting point for research topics as they developed principles for the effective use of regression based on the existing knowledge. They also describe the evidence on the principles. For example, on the matter of evidence on one of the diagnostic statistics they state (p. 311), "The mountains of articles on autocorrelation testing contrast with the near absence of studies on the impact of autocorrelation correction on [*ex ante*] forecast performance."

Unfortunately, software developers typically fail to incorporate evidence-based findings about which methods work best, and statisticians who work on new forecasting procedures seldom cite the literature that tests the relative effectiveness of various methods for forecasting (Fildes and Makridakis 1995). To make the latest findings easily available to forecasters, the International Institute of Forecasters has supported the ForPrin.com website. The goal is to present forecasting principles in a way that can be understood by those who use forecasting techniques such as regression.

S&H recommend visual presentation – and Ziliak (2012) adds support. This seems like a promising avenue. Consider, for example, the effectiveness of the communication provided by Anscombe's Quartet (see Wikipedia).

Conclusions

Do not use regression to search for causal relationships. And do not try to predict by using variables that were not specified in the *a priori* analysis. Thus, avoid data mining, stepwise regression, and related methods.

Regression analysis can play an important role when analyzing non-experimental data. Various illusions can reduce the accuracy of regression analysis, lead to a false sense of confidence, and harm decision-making. Over the past century or so, effective solutions have been developed to deal with the illusions. The basic problem is that the solutions are often ignored in practice. S&H shows that this pertains even among the world's leading researchers. Researchers might benefit by systematically checking their use of regression to ensure that they have taken steps to avoid the illusions. Reviewers could help to make researchers aware of solutions to the illusions. Software providers should inform their users.

To me, S&H's key recommendation is to conduct experiments to compare different approaches to developing and using models. It is remarkable that so little experimentation has been done over the past century to determine which regression methods work best under what conditions.

Arnold Zellner would not have been surprised by these conclusions.

Acknowledgements: Peer review was vital for this paper. Four people reviewed two versions of this paper: P. Geoffrey Allen, Robert Fildes, Kesten C. Green, and Stephen T. Ziliak. Excellent suggestions were also received from Kay A. Armstrong, Jason Dana, Antonio García-Ferrer, Andreas Graefe, Daniel G. Goldstein, Geoff Cumming, Paul Goodwin, Robin Hogarth, Philippe Jacquart, Keith Ord, and Christophe Van den Bulte. This is not to imply that the reviewers agree with all of the conclusions in this paper.

References

- Allen, P. G. & Fildes, R. (2001). Econometric forecasting. In J. S. Armstrong. *Principles of Forecasting*, Boston: Kluwer Academic Publishers,
- Armstrong, J. S. (2011). Evidence-based advertising: An application to persuasion. *International Journal of Advertising*, 30 (5), 743-767.
- Armstrong, J. S. (2001). Judgmental bootstrapping: Inferring experts' rules for forecasting." In J. S. Armstrong. *Principles of Forecasting*. Boston: Kluwer Academic Publishers.
- Armstrong, J. S. (1985). *Long-Range Forecasting*. New York: John Wiley.
- Armstrong, J. S. (1970). How to avoid exploratory research. *Journal of Advertising Research*, 10, No. 4, 27-30.
- Armstrong, J. S. (1968a). *Long-range Forecasting for a Consumer Durable in an International Market*. PhD Thesis: MIT. (URL to be added)
- Armstrong, J. S. (1968b). Long-range forecasting for international markets: The use of causal models. In R. L. King, *Marketing and the New Science of Planning*. Chicago: American Marketing Association. Available at <http://goo.gl/oKYUH>
- Armstrong, J. S. & Collopy, F. (1992). Error measures for generalizing about forecasting methods: Empirical comparisons, *International Journal of Forecasting*, 8, 69-80.
- Armstrong, J. S. & Graefe, A. (2011). Predicting elections from biographical information about candidates: A test of the index method. *Journal of Business Research*, 64, 699-706.
- Armstrong, J. S. & Patnaik, S. (2009). Using Quasi-experimental data to develop principles for persuasive

- advertising. *Journal of Advertising Research*, 49, No. 2, 170-175.
- Christ, C. F. (1960). Simultaneous equation estimation: Any verdict yet? *Econometrica*, 28, 835-845.
- Cumming, G. (2012). *Understanding the New Statistics: Effect sizes. Confidence Intervals and Meta-Analysis*. New York: Routledge.
- Dana, J. & Dawes, R. M. (2004). The superiority of simple alternatives to regression for social science predictions. *Journal of Educational and Behavioral Statistics*, 29 (3), 317-331.
- Einhorn, H. J. (1972). Alchemy in the behavioral sciences. *Public Opinion Quarterly*, 36, 367-378.
- Einhorn, H. J. & R. Hogarth (1975). Unit weighting schemes for decision making. *Organizational Behavior and Human Performance*, 13, 171-192.
- Fildes, R. & Makridakis, S. (1995). The impact of empirical accuracy papers on time series analysis and forecasting. *International Statistical Review*, 63, 289-308.
- Freedman, D. A. (1991). Statistical models and shoe leather. *Sociological Methodology*, 21, 291-313.
- Friedman, M.A. & Schwartz, A. J. (1991). Alternative approaches to analyzing economic data. *American Economic Review*, 81, Appendix 48-49.
- Goldstein, D. G. & Gigerenzer, G. (2009). Fast and frugal forecasting. *International Journal of Forecasting*, 25, 760-772.
- Graefe, A., Armstrong, J.S., Cuzán, A. G. & Jones, R.J., Jr. (2011). Combining forecasts: An application to election forecasts, Working Paper.
- Grove, W.M. & Meehl, P.E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: the clinical – statistical controversy. *Psychology, Public Policy, and Law*, 2, 293-323.
- Hauer, E. (2004). The harm done by tests of significance. *Accident Analysis and Prevention*, 36, 495-500.
- Hogarth, R. M. (2012). When simple is hard to accept. In P. M. Todd & Gigerenzer, G. (Eds.), *Ecological rationality: Intelligence in the world (in press)*. Oxford: Oxford University Press.
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, 2, 696-701.
- Ioannidis, J. P. A. (2008). Why most discovered true associations are inflated. *Epidemiology*, 19, 640-648.
- Karni, E. & Shapiro, B. K. (1980). Tales of horror from ivory towers. *Journal of Political Economy*. 88, No. 1, 210-212.
- Kennedy, P. (2002). Sinning in the basement: What are the rules? The ten commandments of applied econometrics. *Journal of Economic Surveys*, 16, 569-589.
- Meehl, P.E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, 46, 806-834.
- Ord, K. (2012). The Illusion of predictability: A call to action. *International Journal of Forecasting*, 28 xxx-xxx.
- Pant, P. N. & Starbuck, W. H. (1990). Innocents in the forest: Forecasting and research methods. *Journal of Management*, 16, 433-446
- Schmidt, F. L. (1971). The relative efficiency of regression and simple unit predictor weights in applied differential psychology. *Educational and Psychological Measurement*, 31, 699-714.
- Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychological Methods*, 1, 115-129.
- Soyer, E. & Hogarth, R. (2012). Illusion of predictability: How regressions statistics mislead experts. *International Journal of Forecasting*, 28 xxx-xxx.
- Sparks, J. (1844). *The Works of Benjamin Franklin*. Boston: Charles Tappan Publisher.
- Wold, H. & Jureen, L. (1953). *Demand Analysis*. New York: John Wiley.
- Zellner, A. (2001). Keep it sophisticatedly simple. In Keuzenkamp, H. & McAleer, M. Eds. *Simplicity, Inference, and Modelling: Keeping it Sophisticatedly Simple*. Cambridge University Press, Cambridge.
- Ziliak, S. T. (2012). Visualizing uncertainty: On Soyer's and Hogarth's "The illusion of predictability: How regression statistics mislead experts." *International Journal of Forecasting*, 28 xxxxx
- Ziliak, S. T. & McCloskey, D. N. (2008). *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives*. Ann Arbor: The University of Michigan Press.

November 20, 2011 (R47)